

NOTE ON THE MEASUREMENT OF FORECAST SKILL USING A MOVING CLASS INTERVAL

DONALD L. JORGENSEN

Weather Bureau Airport Station, San Francisco, Calif.

[Manuscript received October 13, 1952; revised February 14, 1955]

ABSTRACT

The use of a moving class interval in calculating a skill score for a series of forecasts is investigated. When forecasts are expressed in quantitatively exact terms, it is found that the use of this type of interval in calculating the correct and expected correct number of forecasts gives a good measure of the skill of the forecasting procedure.

1. INTRODUCTION

In the verification of many types of forecasts, it has become common practice to express the data in the form of a contingency table and from this to derive a skill score [1]. The most common type of table consists of two categories, but as many categories as desired can be used. As pointed out by Vernon [2], the conventional skill score gives the same weight to each incorrect forecast regardless of the size of the error. In the same way, each correct forecast is given the same weight regardless of the size of the error, where the size of the greatest possible error which can be verified as correct is a function of the size of the class interval. This latter characteristic becomes of significance when forecasts are issued in quantitative terms, such as temperature forecasts to the exact degree, and then verified by classes. For example, a class interval of 7° would allow a temperature forecast falling at one end of the class interval to be verified by an observed value falling at the other end, even though exhibiting an error of 6° . At the other extreme, a forecast may be evaluated as incorrect with an error as small as 1° for those rather frequent occasions when the forecast value falls in one class interval and the observed value falls just across the class boundary into an adjoining class. In order to compensate for these recognized deficiencies, Vernon has proposed two types of scores to evaluate forecasts in terms of the magnitude of the error (see [2]).

In this paper, a modification of the conventional skill score is proposed which corrects some of the deficiencies resulting from rigid class intervals, while at the same time maintaining the general characteristics of the conventional score.

2. CONCEPT OF A MOVING CLASS INTERVAL

For verification purposes, the usefulness of a series of forecasts may be considered as dependent on the degree to which the verifying observations come within a specified allowable error of the forecast values. The allowable

error depends on the use to which the forecasts are to be put, and in the case of temperature forecasts, for example, might vary from 2° for frost protection when temperatures are near or below freezing to much larger values for general public use. Thus, a verification scheme which evaluates as correct only those forecasts which come within a given plus or minus amount of the observed value has a usefulness well founded in practice. If we assume that a temperature forecast must come within 3° to be of full use to a particular recipient, then the verifying range in which the observed value must fall is within $\pm 3^\circ$ of the forecast value. We see at once that we have what corresponds to a moving class interval with the forecast value at the center and the interval covering a 7° range. A system which uses a moving class interval for forecast verification would verify as correct many forecasts in common with a system using comparable fixed class intervals. However, certain forecasts would verify as correct in one system but not in the other.

3. CONVENTIONAL SKILL SCORE

The conventional skill score S is usually expressed by the relationship:

$$S = \frac{C - E_c}{T - E_c} \quad (1)$$

where

C = number of correct forecasts based on a specified criterion.

T = total number of forecasts.

E_c = number of forecasts which could be expected to be correct on a chance basis with a random distribution of the forecasts over the period for which the score is calculated.

In order to evaluate (1) in terms of a moving class interval, certain of the quantities involved in the expression for the score, i. e., the value for C and E_c , need to be derived in a somewhat different manner, since they are not obtainable from the usual contingency table. The

number of correct forecasts, C , may be obtained by counting the forecasts which fulfill the verifying criterion set forth. In determining the value of E_c , it may be assumed that the probability of an individual forecast to succeed on chance is equal to the ratio of the number of ways it can succeed to the number of ways it can both succeed and fail. Thus, the expression for obtaining the total number of expected correct forecasts for a particular temperature, t , may be written:

$$(E_c)_t = N_t \frac{f_{t \pm 3}}{T} \quad (2)$$

where

N_t = number of times the forecast of the given temperature, t , is made.

$f_{t \pm 3}$ = number of occurrences in array of data of the observed temperatures which fall within $\pm 3^\circ$ of the forecast temperature.

T = total number of cases.

Summing over all forecast temperatures, we may write:

$$E_c = \sum_{t=t'}^{t''} N_t \frac{f_{t \pm 3}}{T} \quad (3)$$

where

t' = lowest forecast temperature.

t'' = highest forecast temperature.

Substituting the quantities C , T , and E_c , thus obtained, into (1) gives the skill score.

4. EXAMPLE OF SKILL SCORE BASED ON MOVING CLASS INTERVAL

Harman [3], in a paper on forecasting maximum temperatures for Sacramento, Calif., verified his forecasts on the basis of a $\pm 3^\circ$ allowable error. Data taken from his study are used to illustrate the calculation of a skill score based on a moving class interval in the following example. In his paper, maximum temperature forecasts for the afternoon are made on data available at 7:15 a. m., local time, and cover a period for July and August for the years 1946 through 1952, with the years of 1951 and 1952 used as test data. Combined dependent and independent data will be used in the example that follows.

The pertinent data for the calculation of a skill score based on a moving class interval appear in table 1. Columns 1 through 4 pertain to observed data, and include the frequency of occurrence of given temperatures, the number of occurrences within $\pm 3^\circ$ of the given temperature, and the ratio of the number of occurrences within $\pm 3^\circ$ to the total number of cases. This latter value is the probability that a forecast of the given temperature will be correct on chance. (These probability values may be calculated on a climatological basis, and once this is done for a particular forecast problem, the calculation of the skill score is simplified.) Column 5 gives the number of forecasts for each temperature, and column 6 the product of the probability ratio in column 4 and the number of

TABLE 1.—Data used in calculating skill score based on a moving class interval of $\pm 3^\circ F.$, Sacramento, Calif., July and August 1946–52

(1) Maximum temperature (°F.)	(2) Frequency of occur- rence of given temp.	(3) Number of occurrences within $\pm 3^\circ$ of given temp.	(4) Ratio of number in col. 3 to total num- ber fore- casts	(5) Number forecasts of given temp.	(6) Product of cols. 4 and 5
	f_t	$f_{t \pm 3}$	$f_{t \pm 3}/T$	N_t	$(E_c)_t$
69	1	3	0.007		
70	1	3	.007		
71	0	5	.012		
72	1	8	.019		
73	0	8	.019	1	0.02
74	2	14	.034		
75	3	18	.043	2	.09
76	1	25	.060	2	.12
77	7	31	.075	2	.15
78	4	43	.104	2	.21
79	8	49	.118	3	.35
80	6	58	.140	5	.70
81	14	59	.143	8	.43
82	9	68	.164	15	2.46
83	10	83	.200	8	1.60
84	8	100	.242	11	2.66
85	13	113	.273	18	4.91
86	23	119	.287	20	5.74
87	23	130	.314	27	8.48
88	27	138	.333	22	7.33
89	15	157	.379	20	7.58
90	21	153	.370	25	9.25
91	16	150	.362	21	7.60
92	32	150	.362	35	12.67
93	19	161	.389	33	12.84
94	20	154	.372	19	7.07
95	27	152	.367	21	7.71
96	26	126	.304	19	5.78
97	14	122	.295	15	4.42
98	14	106	.256	21	5.38
99	6	87	.210	8	1.68
100	15	65	.157	6	.94
101	4	55	.133	4	.53
102	8	43	.104	10	1.04
103	4	38	.092	3	.28
104	4	26	.063	3	.19
105	2	23	.056	3	.17
106	1	16	.039	4	.16
107	3	12	.029	2	.06
108	1	8	.019	1	.02
109	1	6	.014		
Total				414	120.62

forecasts issued in column 5, which furnishes the number of expected correct forecasts for each temperature. The sum of column 6 gives E_c , the number of forecasts expected to be correct on a chance basis.

The number of correct forecasts C , i. e., the number for which the observed temperature fell within $\pm 3^\circ$ of the forecast value, is 310. The total number of forecasts T is 414, and the expected correct, E_c , as given at the bottom of column 6 is 120.6. Substituting these data into (1) gives a skill score of .65, the skill measured by the moving class interval technique.

5. COMPARISON OF SKILL SCORES

The magnitude of the skill score exhibited by a series of forecasts depends upon the particular procedure for calculating the skill. Each method of calculating skill will produce a different figure, with each score representing a somewhat different basis for the evaluation of the forecasts. In as much as the conventional type of score has come to represent a standard, it is of interest to compare the score obtained by the moving class interval with the fixed class interval. The skill score for the same data arranged in a contingency table made up of fixed class intervals of 7° is as follows:

$$S = \frac{276 - 115.5}{414 - 115.5} = .54$$

On the basis of a uniform $\pm 3^\circ$ allowable error, the forecasts verified by means of the moving class interval are correctly verified in every instance in contrast to the case of fixed intervals where errors of as much as 6° can be evaluated as correct and errors of as little as 1° can be evaluated as incorrect. A comparison of the two types of scores indicates the difference in the apparent skill which is gained or lost when the verification is carried out under the differing criteria. Comparing the data contained in the two scores, it is seen that there is a net gain of 34 forecasts evaluated as correct by the moving class interval.

From the standpoint of the user who receives the forecasts and the forecaster who prepares them, the increase in percentage correct from 67 percent to 75 percent as well

as the increase in the skill from .54 to .65 represents a significant improvement in the verifying procedure.

REFERENCES

1. G. W. Brier and R. A. Allen, "Verification of Weather Forecasts," *Compendium of Meteorology*, American Meteorological Society, Boston, 1951, pp. 841-848.
2. E. M. Vernon, "A New Concept of Skill Score for Rating Quantitative Forecasts," *Monthly Weather Review*, vol. 81, No. 10, October 1953, pp. 326-329.
3. W. E. Harman, "An Objective Method for Forecasting the Maximum Temperature at Sacramento, California," U. S. Weather Bureau Office, Pomona, Calif., September 1952 (unpublished manuscript, revised November 1953, available at U. S. Weather Bureau, Washington, D. C.).